DOCUMENT RESUME

ED 364 605 TM 020 877

AUTHOR House, J. Daniel; Keeley, Edward J.
TITLE Differential Prediction of Graduate Student Achievement from Miller Analogies Test Scores.
PUB DATE Nov 93
NOTE 32p.; Paper presented at the Annual Meeting of the Illinois Association for Institutional Research (Oakbrook Terrace, IL, November 4-5, 1993).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Academic Achievement; Admission (School); *Age Differences; Education Majors; Grade Point Average; Grades (Scholastic); *Graduate Students; Higher Education; Masters Programs; *Predictive Validity; *Sex Differences; Test Bias; *Test Validity
IDENTIFIERS *Miller Analogies Test

ABSTRACT

Although the Miller Analogies Test (MAT) is used extensively as an admissions test for graduate programs, relatively little research has examined the validity of the MAT for predicting subsequent graduate student achievement and no published studies have examined differential prediction of student performance from MAT scores. There were two purposes of this study. The first purpose was to investigate the validity of the MAT for predicting the achievement of graduate students in education. The second purpose of this study was to examine the extent of age and gender differences in prediction of graduate student performance from MAT scores. The students included in this analysis were a sample of 1,438 students who began master's level study in education at a large private university during a seven-year period. Results from this study indicated that MAT scores were significantly correlated with students' cumulative grade point averages and with their grade performance in five specific graduate courses. These results also indicated that MAT scores do not predict graduate performance similarly for male and female students and for students of all ages. (Contains 69 references.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 364 605

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
J. DANIEL HOUSE
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Differential Prediction of Graduate Student Achievement From Miller Analogies Test Scores

J. Daniel House
Northern Illinois University

Edward J. Keeley
Loyola University Chicago

Paper presented at the Illinois Association for Institutional Research Annual Meeting, Oakbrook Terrace, IL, November 4-5, 1993.

TM020877

BEST COPY AVAILABLE

Abstract

Although the Miller Analogies Test (MAT) is used extensively as an admissions test for graduate programs, relatively little research has examined the validity of the MAT for predicting subsequent graduate student achievement and no published studies have examined differential prediction of student performance from MAT scores. There were two purposes of this study. The first purpose was to investigate the validity of the MAT for predicting the achievement of graduate students in education. The second purpose of this study was to examine the extent of age and gender differences in prediction of graduate student performance from MAT scores. The students included in this analysis were a sample of 1,438 students who began master's-level study in education at a large private university during a seven-year period. Results from this study indicated that MAT scores were significantly correlated with students' cumulative graduate grade point average and with their grade performance in five specific graduate courses. These results also indicated that MAT scores do not predict graduate performance similarly for male and female students and for students of all ages.

One facet of educational measurement that is of interest to professionals in testing, admissions, and institutional research is the utility of tests used for admissions purposes. For undergraduate admissions, the two most commonly used tests are the American College Testing Program (ACT) and the Scholastic Aptitude Test, which will soon be renamed the Scholastic Assessment Tests (SAT). At the graduate level, the Graduate Record Examination (GRE) and the Miller Analogies Test (MAT) are widely used. Although several previous studies have evaluated the validity of these tests as predictors of subsequent student achievement, it has been recommended that institutions conduct additional research on the predictive validity of admissions tests (Whitney, 1989). These validity studies should investigate the relationship between admissions test scores and several criterion measures of student achievement such as cumulative grade point average (GPA), grades in specific courses, or degree attainment (Hartnett & Willingham, 1980). Despite a large number of previous validity studies, however, considerably fewer studies have investigated whether these tests predict the performance of all subgroups of students similarly (Drasgow, 1987).

Although the MAT is used extensively as an admissions test for graduate programs, relatively little research has examined the validity of the MAT for predicting subsequent graduate student achievement. The small number of existing studies have been characterized by small sample sizes and have produced conflicting results. In fact, no previous studies have investigated whether MAT scores predict subsequent achievement similarly for all subgroups of graduate students. In the following discussion, prior research on the predictive validity of the MAT will be reviewed

and an introduction to differential prediction and prediction bias will be presented. Previous findings on age and gender bias in prediction of student achievement from ACT, SAT, and GRE scores will be reviewed. It will become evident that further study on the predictive validity of the MAT is needed and that research on prediction bias from MAT scores should be conducted.

Predictive Validity of the MAT

A limited number of studies have evaluated the predictive validity of the MAT and those studies have yielded conflicting results. The criterion measure that has been most commonly used in MAT validity studies is cumulative graduate grade point average (GPA). Several studies have found MAT scores to be significant predictors of subsequent grade performance. For example, Furst and Roelfs (1979) found a significant correlation between MAT scores and the graduate GPA of 348 students in a doctoral program in education. Similarly, Cantwell (1990) found that MAT scores were significantly related to the graduate grade performance of 81 students in a counselor education program. Graham (1991) noted a significant relationship between MAT scores and the graduate grades of 50 students in a Master of Business Administration program. Finally, MAT scores were significant predictors of graduate GPA for a sample of 73 students in a master's-level psychology program (Littlepage, Bragg, & Rust, 1978). However, a number of studies have failed to find a significant relationship between MAT scores and graduate GPA. Williams, Harlow, and Gab (1970) evaluated the records of 84 doctoral students in education and found a slightly negative correlation (-.03) between MAT scores and students'

cumulative graduate GPA. Similarly, Federici and Schuerger (1974) failed to find a significant relationship between MAT scores and the graduate grades of students in psychology while Vacc and Picot (1984) also found that MAT scores were not significant predictors of the graduate grades of students in education.

A smaller number of studies have assessed the predictive validity of the MAT for other types of academic outcomes and the results have tended to find MAT scores to be significant predictors of later performance. Two studies have examined students' grades in specific courses as the criterion measure. Mehrabian (1969) found that MAT scores were significantly correlated with grades in specific psychology courses and with statistics grades for graduate students in psychology. Furst and Roelfs (1979) also found that MAT scores were significant predictors of grades in statistics for a sample of graduate students in education. In addition, Kirnan and Geisinger (1981) found that MAT scores were significantly correlated with the performance of graduate students in psychology on their master's comprehensive examinations. Yet another criterion measure that has been studied is degree completion. Three studies have evaluated the predictive validity of the MAT when using the criterion measure of students' completion vs. non-completion of their graduate degrees. None of the three studies found MAT scores to be significant predictors of whether or not students completed their graduate degrees in education (Nagi, 1975; Williams, Harlow, & Gab, 1970) or in psychology (Rawls, Rawls, & Harrison, 1969).

Differential Validity and Prediction Bias

A great deal of research has been directed to the definition and identification of test bias and several approaches have been

proposed. Most interest has been focused on the differential performance of various groups of students, particularly minority students and women (Cole & Moss, 1989; Moore, 1989). There have been numerous studies which have found that minority students attained lower mean scores on standardized tests than majority students (Moore, 1989; Stanley, 1971). Similarly, several studies have found that female students scored lower on specific types of tests, particularly tests of mathematics achievement and specific types of mathematics test items, than did male students (Bridgeman & Wendler, 1991; Doolittle & Cleary, 1987). In addition, considerable attention has been given to the effects of differential test scores on the selection of students for particular programs (such as graduate programs) or on the placement of students into remedial programs (Bickel, Hammel, & O'Connell, 1975; O'Connor, 1989; Mehrens, 1989). An important consideration in the use of standardized tests for selection purposes is that test scores predict subsequent achievement similarly for all groups of students who take the test. When systematic differences exist in the relationship between the test score and the criterion measure for various subgroups of students, then prediction bias exists (Kaplan, 1985; Reynolds, 1982). Prediction bias refers to a systematic error in predicting the criterion variable for particular groups of students. For example, it is important to determine if MAT scores predict the subsequent achievement of graduate students in education similarly for all subgroups of students. In this case, the presence of significant differences between subgroups of students, such as male and female students or younger and older students, in how accurately achievement was predicted from MAT scores would provide evidence of prediction bias.

Two methods have been proposed for the study of bias in prediction of student achievement from admissions test scores. The first approach involves the comparison of validity coefficients for all subgroups of students, and is sometimes referred to as an examination of differential validity (Jensen, 1980). For example, the correlations between MAT scores and graduate GPA would be computed separately for male and female students and those correlations would then be tested for a significant difference. Similarly, age differences in predictive validity would be investigated by computing the validity coefficients for younger and older students and then testing those correlations for a significant difference. Specific statistical tests are available to test for a significant difference between two correlation coefficients (Kleinbaum & Kupper, 1978) or for a significant difference between three or more correlations (Edwards, 1984) and those methods have been applied in previous research on differential validity (House & Johnson, 1992, 1993b).

A second method used to investigate prediction bias involves the comparison of subgroups of students on the mean error of prediction of GPA from admissions test scores (Reynolds, 1982). This approach involves several steps. First, a regression equation based on the entire sample of students is used to compute a predicted GPA for each student. A residual score (the predicted GPA minus the actual GPA) is then computed for each individual. The mean error for each group of students (gender group or age group) is the mean of the residual scores for each individual in the group (Pedhazur, 1982). For example, the mean error of prediction for male students would be the mean of the residual scores of each male student. Finally, subgroups of students are compared using analysis of

variance (ANOVA) to test for significant differences in residual scores, which represents systematic error in the prediction of GPA from admissions test scores. This approach for investigating prediction bias was first developed by Dr. T.A. Cleary and is sometimes referred to as the Cleary model (Hulin, Drasgow, & Parsons, 1983).

Previous Research on Age Bias

In most testing situations, tests are designed for use with younger traditional-aged students. As a result, relatively little is known about the predictive validity of admissions tests for older nontraditional students (Stricker & Rock, 1987). In addition, only a limited number of studies have investigated age differences in the prediction of grade performance from admissions test scores. Results from a study of graduate students in several disciplines indicated that older female students tended to earn graduate grade point averages that were higher than predicted by their GRE scores while younger female students earned graduate grade point averages that were somewhat lower than predicted by their GRE scores (Swinton, 1987). These results provided evidence that GRE scores did not predict subsequent academic performance similarly for students of all ages. However, Swinton (1987) did not evaluate students in graduate programs in education. House (1989) investigated age bias in prediction of graduate grades from GRE scores for a large sample of graduate students in education and obtained results similar to those of Swinton (1987); GRE-Quantitative and GRE-Total scores (GRE-Verbal scores plus GRE-Quantitative scores) underpredicted the performance of older students and overpredicted the graduate grade point average of younger students. It should be noted, however, that despite the presence of age bias in prediction of

grade performance from GRE scores, there was no evidence of differential prediction of achievement from GRE scores. That is, no significant differences were found between younger and older students for the correlations between GRE scores and later graduate grade point average (House, 1989).

Evidence for age bias in prediction of student achievement from admissions test scores has also been reported in a cross-cultural setting. Zeidner (1987b) examined the relationship between SAT scores and academic achievement for students at a university in Israel. One finding was that the correlation between SAT scores and subsequent GPA was significantly lower for students over 30 years of age than for younger students. Zeidner (1987b) also found age bias in the prediction of subsequent grade performance from SAT scores. Finally, age bias in prediction of student grade point average from GRE scores has been reported for a second sample of graduate students in education (Matthews & Martin, 1992).

The findings of these studies indicate that admissions test scores may not predict subsequent GPA similarly for students of all age groups. Student age is a variable that should be considered in future validity studies (Matthews & Martin, 1992). It should also be noted that none of the previous age bias studies examined the MAT as the admissions test. Consequently, further study is needed to determine if similar results would be obtained from the prediction of graduate GPA from MAT scores.

Previous Research on Gender Bias

Several studies have investigated gender bias in prediction of student grade performance from admissions test scores. In general, the results of these studies have indicated that admissions test

scores tend to underpredict the academic performance of female students and to overpredict the achievement of male students. That is, female students tend to earn grade point averages that are higher than were predicted from their test scores while male students tend to earn lower grades than were predicted from their admissions test scores. Research on undergraduate students has found this pattern of underprediction of female students' GPA from both ACT and SAT scores. Considering students at several universities, Linn (1973) found a systematic underprediction of female students' GPA from SAT scores. Breland and Griswold (1982) also found that female students' grade point averages were underestimated by ACT scores while Zeidner (1987a) provided similar evidence of gender bias in grade performance from SAT scores for students in a cross-cultural setting. Finally, Sawyer (1986) examined gender bias in prediction of first-year college grades from ACT scores using data from over 100,000 students at over 200 colleges. As was the case with findings from the SAT, Sawyer (1986) found that ACT scores underpredicted the subsequent grade performance of female students and overpredicted the first-year college grades of male students.

One factor that may influence gender bias in prediction of grade performance from admissions test scores is that male and female students may enroll in different types of courses that might have differing grading standards. Previous research has suggested that there are considerable differences between major fields in the grading standards that are used (Goldman, Schmidt, Hewitt, & Fisher, 1974; Strenta & Elliot, 1987) while other studies have shown that the predictive relationship between admissions test scores and grade performance can differ substantially within distinct program areas in in the same discipline (House & Johnson, 1993a; Kirnan & Geisinger,

1981). Consequently, research that considered the effects of different majors for male and female students has provided additional insight into the extent of gender bias. Gamache and Novick (1985) attempted to control for the effects of differential grading by separating students by their major fields: Business, Liberal Arts, Pre-Medicine, and Undecided. Their findings indicated that differential prediction was evident within programs and that, as was found in earlier research, ACT scores underpredicted the actual grade performance of female students and overpredicted the grade point average of male students. Finally, McCornack and McLeod (1988) controlled for the effects of differing grading standards by examining course grades as the criterion measure. In contrast to previous findings, McCornack and McLeod (1988) failed to find evidence of gender bias in the prediction of college course grades from SAT scores. Consequently, the appearance of gender bias in the prediction of GPA from admissions test scores may be due to the effects of differential grading practices on the criterion measure of cumulative GPA.

There has been little research on gender bias in the prediction of graduate student performance from GRE scores. Kaczmarek and Franco (1986) found that GRE scores did not predict subsequent grade performance similarly for a small sample of male and female students in a master's-level program in counseling. There have been no studies that have examined gender bias in the prediction of graduate student performance from MAT scores. Consequently, research is needed to investigate gender bias in prediction from MAT scores, and that research should consider the effects of differential grading standards on gender bias.

Statement of Purpose

There were three main purposes of this study. First, this study was designed to investigate the validity of the MAT for the prediction of subsequent graduate student performance. Because previous research has indicated that the predictive validity of an admissions test can vary substantially for different types of criterion measures (Hartnett & Willingham, 1980; House, Johnson, & Tolone, 1987), several criterion measures of academic performance were examined. The second purpose of this study was to investigate the extent of gender bias in the prediction of student achievement from MAT scores. Considerable research has assessed gender bias in the prediction of undergraduate students' grade performance from ACT and SAT scores. This study was designed to build upon those studies by examining gender bias in the prediction of graduate student achievement. Finally, the third purpose of this study was to examine the extent of age bias in the prediction of graduate student performance from MAT scores. Previous research on age bias has been conducted using GRE scores as predictors (House, 1989; Swinton, 1987), and this study was designed to extent those findings.

Methods

Students

The students included in this analysis were a sample of 1,438 students (mean age = 32.7 years, SD = 8.00) who began master's-level study in education at a private university during a seven-year period. In this sample, there were 192 male students and 1,246 female students. In the entire sample, there 75 African-American students, 15 Asian students, 29 Hispanic students, and 1,306 white students; there were 13 students in this sample whose ethnic status

was identified as other or not given. For analyses done by age group, there were 261 students who were 24 years of age or younger when they began graduate study and there were 1,177 students who were 25 years of age or older when they began graduate study.

Measures

Demographic data collected for each student included their gender, ethnic group, and age at the time they began graduate study. Each students' MAT score was also collected. Finally, data for eight criterion measures of academic performance were also compiled for each student. Two measures related to degree attainment were collected: whether or not students completed their master's degree and the length of time (in years) required for the completion of the degree. Finally, five measures of grade performance were also collected: students' cumulative graduate grade point average and their grades in five courses required of all graduate students in education (Survey of Child Development, Theories of Teaching and Learning, Introduction to Graduate Research, Research for Teachers, and History and Philosophy of Education).

Procedures

The data from this study were analyzed in several ways. First, descriptive statistics for MAT scores and each measure of grade performance were computed. This was done for the entire sample and separately for male and female students and by student age group. In order to be consistent with earlier research of this type, younger students were defined as 24 years of age and younger at the time graduate study was initiated while older students were those who were 25 years of age and older.

In order to investigate the predictive validity of the MAT, Pearson product-moment validity coefficients were computed for the

MAT and each criterion measure. For a criterion measure that is binary (graduated vs. not-graduated), the point-biserial correlation is ordinarily recommended. However, the Pearson correlation coefficient provides a satisfactory estimate of the true correlation between a binary variable and a continuous variable (Leonard, 1976). These correlation coefficients between MAT scores and each criterion measure were computed for the entire sample and separately for male and female students. Correlations were also computed separately for younger and older students. The correlation coefficients for male and female students and for younger and older students were then tested for equivalence using a Z-transformation procedure (Kleinbaum & Kupper, 1978).

The method used to determine prediction bias was to compare students by age group and by gender group on the mean error of prediction found for the MAT. A regression equation based on all students was used to compute a predicted GPA for each student. A residual score (the predicted GPA minus the actual GPA) was then computed for each individual. The mean error for each group was the mean of the residual scores, and the groups were then compared using analysis of variance (ANOVA) to test for systematic error in the prediction of graduate grades from MAT scores. This procedure was done separately for cumulative graduate GPA and for grades in each of the five required graduate courses.

Results

Descriptive Statistics

Descriptive statistics for MAT scores, the length of time required for master's degree completion, and each measure of grade performance for the entire sample are presented in Table 1. The

descriptive statistics for male and female students are also summarized in Table 1. Preliminary tests indicated that, for this sample, male students had significantly higher MAT scores ($t(1436) = 2.17$, $p = .0301$) while female students had significantly higher cumulative graduate grade point averages ($t(1436)=5.57$, $p=.0001$). Female students also earned significantly higer grades than did male students in two of the five specific courses that were examined. In each case, the initial tests for homogeneity of variances indicated that male and female students had significantly different variances for their course grades; consequently, a t-test procedure for unequal variances was used (Leonard, 1976). Female students

Table 1

Descriptive Statistics for Predictor Variables and Criterion Measures (For All Students and by Student Gender)

	All		Males		Females	
Variable	M	SD	M	SD	M	SD
Miller Analogies Test	47.67	15.12	50.02	16.33	47.31	16.07
Cumulative Graduate Grade Point Average	3.83	0.24	3.74	0.35	3.85	0.22
Time to Degree (Years)	2.82	1.48	2.56	1.32	2.86	1.50
Survey of Child Development	3.84	0.39	3.70	0.55	3.86	0.36
Theories of Teaching and Learning	3.77	0.46	3.68	0.55	3.78	0.44
Introduction to Graduate Research	3.80	0.43	3.72	0.53	3.82	0.41
Research for Teachers	3.88	0.35	3.84	0.38	3.88	0.34
History & Philosophy of Education	3.78	0.46	3.73	0.53	3.78	0.45

earned significantly higher grades in Survey of Child Development ($t(100.9) = 2.75$, $p = .0007$) and in Theories of Teaching and Learning ($t(158.9) = 2.04$, $p = .0428$).

Descriptive statistics summarized by student age group are presented in Table 2. Preliminary tests indicated that older students had significantly higher MAT scores than did younger students ($t(446) = 6.74$, $p = .0001$). Older students had significantly higher grades in Theories of Teaching and Learning ($t(244.7) = 3.10$, $p = .0022$) and in Research for Teachers ($t(255.4) = 3.62$, $p = .0004$). The analyses for each course used the t-test procedure for unequal variances.

Table 2

Descriptive Statistics for Predictor Variables and Criterion Measures (By Student Age Group)

	Younger		Older	
Variable	M	SD	M	SD
Miller Analogies Test	42.34	13.56	48.85	16.41
Cumulative Graduate Grade Point Average	3.81	0.21	3.84	0.25
Time to Degree (Years)	2.78	1.55	2.84	1.47
Survey of Child Development	3.83	0.39	3.84	0.39
Theories of Teaching and Learning	3.66	0.52	3.80	0.44
Introduction to Graduate Research	3.80	0.44	3.81	0.42
Research for Teachers	3.78	0.43	3.90	0.33
History & Philosophy of Education	3.77	0.45	3.78	0.46

Predictive Validity of the MAT

Correlations between MAT scores and each criterion measure of academic performance for the entire sample are summarized in Table 3. MAT scores were significant predictors of students' cumulative graduate GPA and of their grade performance in each of the five specific courses. However, MAT scores were not significantly correlated with whether or not students had completed their master's degrees or with the length of time (in years) required to complete a master's degree in education. The relationship between MAT scores and graduate student acheivement was also examined separately for male and female students and those results are also presented in Table 3. When considering male students, the correlations that were found to be significant were the same as was the case for the entire

Table 3

Correlations Between MAT Scores and Criterion Measures (For All Students and By Student Gender)

Outcome Measure	All Students	Male Students	Female Students	Z
Cumulative Graduate Grade Point Average	.273**	.255**	.298**	0.62
Graduated/Not Graduated	.013	-.075	.027	1.34
Time to Degree Completion	.057	.020	.065*	0.50
Survey of Child Development	.188**	.325**	.180**	1.40
Theories of Teaching & Learning	.241**	.354**	.223**	1.51
Intro to Graduate Research	.256**	.352**	.247**	1.21
Research for Teachers	.167**	.244**	.156**	1.02
History & Philosophy of Educ.	.282**	.252**	.290**	0.42

**$p < .01$, *$p < .05$.

sample; for male students, MAT scores were significantly correlated with subsequent cumulative graduate GPA and with grades in each of the five specific courses. For female students, MAT scores were significantly related to seven of the eight criterion measures; only the correlation between MAT scores and degree completion was not significant. Finally, no significant differences between male and female students were noted for any of the relationships between MAT scores and the academic performance measures included in this study. The lack of any significant differences indicates that the correlations found for male and female students can be considered estimates of the same population values.

Correlations between MAT scores and each measure of academic performance were analyzed separately for each student age group and those findings are presented in Table 4. For younger students, MAT

Table 4

Correlations Between MAT Scores and Criterion Measures (By Student Age Group)

Outcome Measure	Younger (Less Than 24)	Older (25+)	Z
Cumulative Graduate Grade Point Average	.274**	.269**	0.07
Graduated/Not Graduated	.010	.020	0.14
Time to Degree Completion	.020	.063	0.56
Survey of Child Development	.237**	.176**	0.75
Theories of Teaching & Learning	.210**	.234**	0.28
Intro to Graduate Research	.199**	.276**	1.14
Research for Teachers	.097	.163**	0.88
History & Philosophy of Educ.	.273**	.290**	0.25

**$p < .01$, *$p < .05$.

scores were significant predictors of cumulative graduate GPA and of grade performance in four of the specific courses. For older students, MAT scores were significant predictors of cumulative graduate GPA and of grade performance in all five of the specific courses. The relationship between MAT scores and subsequent degree completion was not significant for either group of students. Similarly, the relationship between MAT scores and the length of time (in years) required for students to complete their master's degree was not significant for either group of students. Finally, no significant differences between younger and older students were found for any of the correlations between MAT scores and any of of the eight measures of academic performance.

Prediction Bias

Findings from the analyses of gender bias in prediction of each measure of grade performance are summarized in Table 5 and several

Table 5

Gender Differences in Prediction of Grade Performance From Miller Analogies Test Scores

	Mean Error				
Outcome	Males	Females	F	df	p
Cumulative Graduate Grade Point Average	.099	-.015	41.39	1,1436	.0001
Survey of Child Development	.160	-.020	18.66	1,820	.0001
Theories of Teaching and Learning	.103	-.019	8.07	1,825	.0046
Introduction to Graduate Research	.099	-.014	8.85	1,1085	.0030
Research for Teachers	.042	-.007	2.41	1,989	.1209
History & Philosophy of Education	.067	-.009	3.05	1,965	.0809

significant results were obtained. First, a significant difference between male and female students for the mean error of prediction of cumulative graduate GPA from MAT scores was found. In this case, MAT scores significantly underpredicted the cumulative graduate grade point average of female students and overpredicted the cumulative graduate GPA of male students. This finding indicates that female students earned higher cumulative graduate grade point averages than were predicted from their MAT scores while male students earned lower cumulative graduate grade point averages than were predicted by their MAT scores. Similar findings were also noted for three of the five specific graduate courses. Significant differences between male and female students for the mean error of prediction of grade performance in those three courses were found. For those three courses (Survey of Child Development, Theories of Teaching and Learning, and Introduction to Graduate Research), female students earned grades that were higher than were predicted by their MAT scores while male students earned grades that were lower than were predicted by their MAT scores. Finally, although significant differences between male and female students for the mean error of prediction of grade performance in the remaining two courses (Research for Teachers and History and Philosophy of Education) were not found, the same pattern of underprediction of female grade performance from MAT scores was observed.

Findings from the analyses of age bias in prediction of grade performance from MAT scores are presented in Table 6. There was not a significant difference between younger and older students for the mean error of prediction of cumulative graduate grade point average from MAT scores. This indicates that MAT scores predicted graduate GPA equally accurately for younger and older students. Significant

differences between younger and older students for the mean error of prediction of grade performance in two of the five specific courses were found. For those two courses (Theories of Teaching and Learning and Research for Teachers), MAT scores underpredicted the grades earned by older students and overpredicted the grades earned by younger students. This finding indicates that, for those two courses, older students earned grades that were higher than were predicted by their MAT scores while younger students earned grades that were lower than were predicted by their MAT scores. Finally, for the three courses for which there were not significant differences in the mean error of prediction, MAT scores tended to underpredict the grade performance of younger students and overpredict the grades earned by older students.

Table 6

Age Differences in Prediction of Grade Performance From Miller Analogies Test Scores

	Mean Error				
Outcome	Younger	Older	F	df	p
Cumulative Graduate Grade Point Average	-.001	.000	0.01	1,1436	.9188
Survey of Child Development	-.015	.004	0.38	1,820	.5355
Theories of Teaching and Learning	.071	-.019	5.67	1,825	.0175
Introduction to Graduate Research	-.039	.010	2.57	1,1085	.1094
Research for Teachers	.073	-.018	10.95	1,989	.0010
History and Philosophy of Education	-.045	.012	2.67	1,965	.1023

Discussion

There were three main findings from this study. First, MAT scores were found to be significant predictors of cumulative graduate GPA and of grade performance in five specific graduate courses. These findings are consistent with the results of previous studies that have found significant correlations between MAT scores and cumulative graduate GPA (Cantwell, 1990; Furst & Roelfs, 1979; Graham, 1991) and between MAT scores and grades in specific graduate courses (Furst & Roelfs, 1979; Mehrabian, 1969). There was not a significant relationship between MAT scores and students' degree completion status in this study. This lack of a significant predictive relationship between MAT scores and degree completion is also similar to the findings of previous research (Nagi, 1975; Rawls, Rawls, & Harrison, 1969; Williams, Harlow, & Gab, 1970).

The second finding from this study was that MAT scores did not predict graduate grade performance similarly for male and female students. For cumulative graduate GPA and for grades in three of the five graduate courses examined, MAT scores underpredicted the achievement of female students and overpredicted the achievement of male students. The results for cumulative graduate GPA are similar to previous results from undergraduate admissions tests (Linn, 1973; Sawyer, 1986; Zeidner, 1987). However, because it has been suggested that gender bias should not be apparent for the prediction of grades in specific courses (McCornack & McLeod, 1988), the findings of gender bias for the grades in specific courses indicate that factors other than simply differential grading standards are involved in the appearance of gender bias.

The third finding from this study was that MAT scores did not consistently predict graduate grade performance in a similar manner

for younger and older students. There was not a significant difference between younger and older students in the mean error of prediction of cumulative graduate GPA. This lack of significance differs from previous research that found age bias in the prediction of cumulative graduate GPA from GRE scores (House, 1989). However, for two of the five graduate courses examined in this study, MAT scores underpredicted the grades earned by older students and over-predicted the grades earned by younger students. The pattern of older students earning higher grades than predicted by their test scores is similar to findings from previous studies that used cumulative graduate GPA as the criterion measure (House, 1989; Swinton, 1987; Zeidner, 1987b).

There are some limitations to the present study. First, only students at one institution were included in the study. Further research is needed to determine if these results would be found for students at other types of institutions. Previous multi-institutional validity studies have been conducted. Rubin (1980) conducted an analysis of the predictive validity of the Law School Aptitude Test (LSAT) at 82 law schools while Zwick (1993) evaluated the predictive validity of the Graduate Management Admission Test (GMAT) for students in doctoral program in business and management at 36 universities. Similar research is needed to assess the predictive validity of the MAT for graduate students in education at various types of institutions. A second limitation of the present study is that no analysis of the effects of restriction of range were made. It has been shown that restricting the distribution of admissions test scores for students who have been admitted for graduate study will produce lower than expected predictive validity (Givner & Hynes, 1979). Previous studies have investigated the effects of

restriction of range on the predictive validity of the GRE in an actual admissions situation (House, 1983) and the predictive validity of the GRE when restriction of range was not a factor Huitema & Stein, 1993). Similar studies need to be done to assess the effects of restriction of range on the predictive validity of the MAT found in this study. Finally, based on the recommendations made by Sawyer (1987), there were insufficient numbers of minority students in this sample to allow meaningful analyses of ethnic bias in prediction of grade performance. Studies of ethnic bias have been conducted for the ACT (Maxey & Sawyer, 1981; Sawyer, 1986) and for the SAT (Cameron, 1989). Research is needed to determine if MAT scores predict subsequent grade performance similarly for graduate students of all ethnic groups.

The results of this study suggest a number of directions for further research. First, although age and gender differences were found for the relationship between MAT scores and grade performance, no analysis was made of age or gender differences in performance on individual items of the MAT. A number of techniques have been developed to investigate item bias (Holland & Thayer, 1988; Ironson, 1982; Thissen, Steinberg, & Wainer, 1988). Previous studies of gender differences in performance on individual tests items have been conducted for several admissions tests, including the ACT (Handrick & Loyd, 1982), the SAT (Becker, 1990; Carleton & Marco, 1982), and the GRE (Donlan, Hicks, & Wallmark, 1980). Similar research is needed to determine if item bias would be found for the MAT. A second direction for further research would be to determine if the predictive validity of the MAT is similar for graduate students in a variety of disciplines. Previous research has shown that the predictive validity of the GRE can vary considerably for

different fields of graduate study (Thornell & McCoy, 1985). A final direction for further research would be to investigate the improvement in prediction of graduate student performance by combining MAT scores with noncognitive measures. Previous research on undergraduate students has indicated that noncognitive variables are significant predictors of several types of academic outcomes such as grade performance in specific courses (Gordon, 1989; House, 1993a) and of withdrawal from college (House, 1992b, 1993b). Recent research has indicated that, in some instances, noncognitive variables are more significant than admissions tests scores as predictors of student achievement (House, 1992a). Consequently, research is needed to determine if combining noncognitive measures with MAT scores would result in improved prediction of graduate student performance.

These results indicate that, for some measures of graduate student achievement, the predictive validity of the MAT is significant. The findings of this study also indicate that the MAT does not predict subsequent achievement similarly for male and female students or for students of all ages. Finally, these results provide directions for further research.

References

Becker, B.J. (1990). Item characteristics and gender differences on the SAT-M for mathematically able youths. American Educational Research Journal, 27, 65-87.

Bickel, P.J., Hammel, E.A., & O'Connell, J.W. (1975). Sex bias in graduate admissions: Data from Berkeley. Science, 187, 398-404.

Breland, H.M., & Griswold, P.A. (1982). Use of a performance test as a criterion in a differential validity study. Journal of Educational Psychology, 74, 713-721.

Bridgeman, B., & Wendler, C. (1991). Gender differences in predictors of college mathematics performance and in college mathematics courses grades. Journal of Educational Psychology, 83, 275-284.

Cameron, R.G. (1989). Issues in testing bias. College and University, 64, 269-279.

Cantwell, Z.M. (1990). Predictors of performance in a field-based counselor education program. Psychological Reports, 66, 151-159.

Carleton, S.T., & Marco, G.L. (1982). Methods used by test publishers to "debias" standardized tests: Educational Testing Service. In R.A. Berk (Ed.), Handbook of Methods for Detecting Test Bias (pp. 278-308). Baltimore, MD: Johns Hopkins University Press.

Cole, N.S., & Moss, P.A. (1989). Bias in test use. In R.L. Linn (Ed.), Educational Measurement (3rd. ed.). New York: American Council on Education and Macmillan Publishing Co.

Donlan, T.F., Hicks, M.M., & Wallmark, M.M. (1980). Sex differences in item responses on the Graduate Record Examination. Applied Psychological Measurement, 4, 9-20.

Doolittle, A.E., & Cleary, T.A. (1987). Gender-based differential item performance in mathematics achievement items. Journal of Educational Measurement, 24, 157-166.

Drasgow, F. (1987). Study of the measurement bias of two standardized psychological tests. Journal of Applied Psychology, 72, 19-29.

Edwards, A.L. (1984). An Introduction to Linear Regression and Correlation (2nd. ed.). New York: Freeman.

Federici, L., & Schuerger, J. (1974). Prediction of success in an applied M.A. psychology program. Educational and Psychological Measurement, 34, 945-952.

Furst, E.J., & Roelfs, P.J. (1979). Validation of the Graduate Record Examinations and the Miller Analogies Test in a doctoral program in education. Educational and Psychological Measurement, 39, 147-151.

Gamache, L.M., & Novick, M.R. (1985). Choice of variables and gender differentiated prediction within selected academic programs. Journal of Educational Measurement, 22, 53-70.

Givner, N., & Hynes, K. (1979). Achievement test validity: Correcting for restriction effects. College and University, 54, 119-123.

Goldman, R.D., Schmidt, D.E., Hewitt, B.N., & Fisher, R. (1974). Grading practices in different major fields. American Educational Research Journal, 11, 343-357.

Gordon, R.A. (1989). Intention and expectation measures as predictors of academic performance. Journal of Applied Social Psychology, 19, 405-415.

Graham, L.D. (1991). Predicting academic success of students in a Master of Business Administration program. Educational and Psychological Measurement, 51, 721-727.

Handrick, F.A., & Loyd, B.H. (1982). Methods used by test publishers to "debias" standardized tests: The American College Testing Program. In R.A. Berk (Ed.), Handbook of Methods for Detecting Test Bias (pp. 272-278). Baltimore, MD: Johns Hopkins University Press.

Hartnett, R.T., & Willingham, W.W. (1980). The criterion problem: What measure of success in graduate education? Applied Psychological Measurement, 4, 281-291.

Holland, P.W., & Thayer, D.W. (1988). Differential item performance and the Mantel-Haenszel procedure. In H. Wainer & H.I. Braun (Eds.), Test Validity (pp. 129-145). Hillsdale, NJ: Lawrence Erlbaum Associates.

House, J.D. (1983). Effects of restriction of range on predictive validity for the Graduate Record Examination. Psychological Reports, 53, 710.

House, J.D. (1989). Age bias in prediction of graduate grade point average from Graduate Record Examination scores. Educational and Psychological Measurement, 49, 663-666.

House, J.D. (1992a). Noncognitive variables as predictors of achievement in introductory college chemistry. Paper presented at the Illinois Association for Institutional Research annual meeting, Carbondale, Illinois.

House, J.D. (1992b). The relationship between academic self-concept, achievement-related expectancies, and college attrition. Journal of College Student Development, 33, 5-10.

House, J.D. (1993a). Achievement-related expectancies, academic self-concept, and mathematics performance of academically underprepared adolescent students. Journal of Genetic Psychology, 154, 61-71.

House, J.D. (1993b). The relationship between academic self-concept and school withdrawal. Journal of Social Psychology, 133, 125-127.

House, J.D., & Johnson, J.J. (1992). Predictive validity of Graduate Record Examination scores and undergraduate grades for length of time to completion of degree. Psychological Reports, 71, 1019-1022.

House, J.D., & Johnson, J.J. (1993a). Graduate Record Examination scores and academic background variables as predictors of graduate degree completion. Educational and Psychological Measurement, 53, 551-556.

House, J.D., & Johnson, J.J. (1993b). Predictive validity of the Graduate Record Examination Advanced Psychology Test for graduate grades. Psychological Reports, 73, 184-186.

House, J.D., Johnson, J.J., & Tolone, W.L. (1987). Predictive validity of the Graduate Record Examination for performance in selected graduate psychology courses. Psychological Reports, 60, 107-110.

Huitema, B.E., & Stein, C.R. (1993). Validity of the GRE without restriction of range. Psychological Reports, 72, 123-127.

Hulin, C.L., Drasgow, F., & Parson, C.K. (1983). Item Response Theory: Application to Psychological Measurement. Homewood, IL: Dow Jones-Irwin.

Ironson, G.H. (1982). Use of chi-square and latent trait approaches for detecting item bias. In R.A. Berk (Ed.), Handbook of Methods for Detecting Test Bias (pp. 117-160). Baltimore, MD: Johns Hopkins University Press.

Jensen, A.R. (1980). Bias in Mental Testing. New York, NY: The Free Press.

Kaczmarek, M., & Franco, J.N. (1986). Sex differences in prediction of academic performance by the Graduate Record Examination. Psychological Reports, 59, 1197-1198.

Kaplan, R.M. (1985). The controversy related to the use of psychological tests. In B.B. Wolman (Ed.), Handbook of Intelligence: Theories, Measurement, and Applications (pp. 465-504). New York: John Wiley & Sons, Inc.

Kirnan, J.P., & Geisinger, K.F. (1981). The prediction of graduate school success in psychology. Educational and Psychological Measurement, 41, 815-820.

Kleinbaum, D.G., & Kupper, L.L. (1978). Applied Regression Analysis and Other Multivariable Methods. North Scituate, MA: Duxbury Press.

Leonard, W.M. (1976). Basic Social Statistics. New York: West Publishing Co.

Linn, R.L. (1973). Fair test use in selection. Review of Educational Research, 43, 139-161.

Littlepage, G.E., Bragg, D.M., & Rust, J.O. (1978). Relations between admissions criteria, academic performance, and professional performance. Teaching of Psychology, 5, 16-20.

Matthews, T.A., & Martin, D.J. (1992). Reciprocal suppression and interaction effects of age with undergraduate grades and GRE on graduate performance in a College of Education. Educational and Psychological Measurement, 52, 453-456.

Maxey, J., & Sawyer, R. (1981). Predictive validity of the ACT Assessment for Afro-American/Black, Mexican-American/Chicano, and Caucasian-American/White students. American College Testing Program Research Bulletin No. 81-1. Iowa City, IA: American College Testing Program.

McCornack, R.L., & McLeod, M.M. (1988). Gender bias in the prediction of college course performance. Journal of Educational Measurement, 25, 321-331.

Mehrabian, A. (1969). Undergraduate ability factors in relationship to graduate performance. Educational and Psychological Measurement, 29, 409-419.

Mehrens, W.A. (1989). Using test scores for decision making. In B.R. Gifford (Ed.), Test Policy and Test Performance: Education, Language, and Culture (pp. 93-113). Boston, MA: Kluwer Academic Publishers.

Moore, E.G.J. (1989). Ethnic group differences in the Armed Services Vocational Aptitude Battery (ASVAB) performance of American Youth: Implications for career prospects. In B.R. Gifford (Ed.), Test Policy and Test Performance: Education, Language, and Culture (pp. 183-206). Boston, MA: Kluwer Academic Publishers.

Nagi, J.L. (1975). Predictive validity of the Graduate Record Examination and the Miller Analogies Test. Educational and Psychological Measurement, 35, 471-472.

O'Connor, M.C. (1989). Aspects of differential performance by minorities on standardized tests: Linguistic and sociocultural factors. In B.R. Gifford (Ed.), Test Policy and Test Performance: Education, Language, and Culture (pp. 129-181). Boston, MA: Kluwer Academic Publishers.

Pedhazur, E.J. (1982). Multiple Regression in Behavioral Research (2nd. ed.). New York: Holt, Rinehart, and Winston.

Rawls, J.R., Rawls, D.J., & Harrison, C.W. (1969). An investigation of success predictors in graduate school in psychology. Journal of Psychology, 72, 125-129.

Reynolds, C.R. (1982). Methods for detecting construct and predictive bias. In R.A. Berk (Ed.), Handbook of Methods for Detecting Test Bias (pp. 199-227). Batimore, MD: Johns Hopkins Press.

Rubin, D.B. (1980). Using empirical Bayes techniques in the law school validity studies. Journal of the American Statistical Association, 75, 801-816.

Sawyer, R. (1986). Using demographic subgroup and dummy variable equations to predict college freshman grade average. Journal of Educational Measurement, 23, 131-145.

Saywer, R. (1987). Determining minimum sample sizes for estimating prediction equations for college freshmen grade average. ACT Research Report Series No. 87-4. Iowa City, IA: American College Testing Program.

Stanley, J.C. (1971). Predicting college success of the educationally disadvantages. Science, 171, 640-647.

Strenta, A.C., & Elliot, R. (1987). Differential grading standards revisited. Journal of Educational Measurement, 24, 281-291.

Stricker, L.J., & Rock, D.A. (1987). Factor structure of the GRE General Test in young and middle adulthood. Developmental Psychology, 23, 526-536.

Swinton, S.S. (1987). The predictive validity of the restructured GRE with particular attention to older students. GRE Board Report No. 83-25P. Princeton, NJ: Educational Testing Service.

Thissen, D., Steinberg, L., & Wainer, H. (1988). Use of item response theory in the study of group differences in trace lines. In H. Wainer & H.I. Braun (Eds.), Test Validity (pp. 147-169). Hillsdale, NJ: Lawrence Erlbaum Associates.

Thornell, J.G., & McCoy, A. (1985). The predictive validity of the Graduate Record Examinations for subgroups of students in different academic disciplines. Educational and Psychological Measurement, 45, 415-419.

Vacc, N.N., & Picot, R. (1984). Predicting success in doctoral study. College Student Journal, 18, 113-116.

Whitney, D.R. (1989). Educational admissions and placement. In R.L. Linn (Ed.), Educational Measurement (3rd. ed.). New York: American Council on Education and Macmillan Publishing Co.

Williams, J.D., Harlow, S.D., & Gab, D. (1971). A longitudinal study examining prediction of doctoral success: Grade point average as criterion, or graduation vs. non-graduation as criterion. Journal of Educational Research, 64, 161-164.

Zeidner, M. (1987a). A cross-cultural test of sex bias in the predictive validity of Scholastic Aptitude Examinations: Some Israeli findings. Evaluation and Program Planning, 10, 289-295.

Zeidner, M. (1987b). Age bias in the predictive validity of Scholastic Aptitude Tests: Some Israeli data. Educational and Psychological Measurement, 47, 1037-1047.

Zwick, R. (1993). The validity of the GMAT for the prediction of grades in doctoral study in business and management: An empirical Bayes approach. Journal of Educational Statistics, 18, 91-107.

Acknowledgement: The authors with to thank Dr. Robert Clark and Mr. Robert Myers at National-Louis University for permission to use data gathered from their institution for this study.